Thresholds for Graph Properties and a Proof of the Kahn-Kalai Conjecture

Candidate Number: 23890

A Dissertation submitted to the Department of Mathematics of the London School of Economics and Political Science

March 16, 2024

Summary

In this dissertation we examine the result of Park and Pham (2022) which introduces a randomized iterative process to prove the Kahn-Kalai conjecture. We set up the framework for randomized algorithms on discrete probability spaces and study the notion of threshold functions for monotone properties. The distinction between Erdős-Rényi thresholds and a more stringent definition used in the above result, as well as the proof of the Bollobás-Thomason theorem are presented as building blocks.

We illustrate the strength of Park and Pham's result with the problem of finding the threshold function for Squares of Hamilton Cycles. Finally, we carefully show the reformulation of the original conjecture and explain the proof elucidated in the paper above. We pay particular attention to three cornerstones of the proof: exponentially decreasing concentration bounds, expectation thresholds and 'minimum fragments'.

Contents

Chapter 1

Introduction

With their seminal publication 'On the Evolution of Random Graphs' Erdős and Rényi (1960), Erdős and Rényi in 1959 mark the premises of an active area of research at the confluence of Graph Theory and Probability: the study of Random Graphs. A graph is a mathematical object defined by a pair of two sets: one set is composed of vertices, and the other consists of some subset of the collection of all possible edges between these vertices. A random graph is a graph with a stochastic set of edges. Informally, it is the result of, given the set of vertices, drawing each edge uniformly.

In their investigation of the five characteristic phases which compose the evolution of a Random Graph, the authors refer to a 'most surprising fact discovered': when the number of edges passes a certain threshold, the structure of the graph changes abruptly. It undergoes a sudden transition from a state where, almost surely, most components are either trees or components containing exactly one cycle, and the greatest component is a tree with $O(\log n - \frac{5}{2}\log\log n)$ vertices, to a state where the greatest component has $O(n^{\frac{2}{3}})$ vertices and has a complex structure (Bollobás, Janson, and Riordan (2007)).

Considerable attention has been directed to the body of methodologies which provides tools to understand fluctuating properties of random graphs. These are of particular interest as they can be used to model real-world networks such as power grids, social networks and the Internet when size and complexity renders analytical approaches computationally intractable van der Hofstad (2017). In particular, under assumptions of non-triviality and monotonicity, every property admits such a threshold function Bollobás and Thomason (1987).

This has the two-fold striking implication: arbitrarily close to the threshold but below, a random graph will almost surely (a.s.) not possess the property, but crossing the threshold ensures the emergence of the property a.s.. Finding bounds on these functions has been the subject of much attention in this field (e.g. Johansson, Kahn, and Vu (2008), Kahn and Kalai (2007), Bell and Frieze (2024)).

In 2007, Kahn and Kalai (Kahn and Kalai (2007)) make the following powerful conjecture regarding the location of these thresholds – which we later reiterate formally: given a Random Graph on n vertices and for all monotonous, non trivial properties, there exists a naive lower bound on the threshold, the 'Expectation Threshold', which is always within a logarithmic factor – $O(\log n)$ of the truth. To underline the bold nature of this universal claim, the authors state that it would 'be more sensible to conjecture that it is not true'.

However, in 2022, Park and Pham prove this conjecture. Circumventing the crucial use of 'spread' leveraged in preceding breakthroughs regarding the Sunflower Conjecture by Alweiss et. al. Alweiss, Lovett, Wu, and Zhang (2021) and the proof the 'Fractional Kahn-Kalai Conjecture' Frankston, Kahn, Narayanan, and Park (2021), the crux of this result lies in construction of a 'cheap' cover for a property through a random iterative process.

In this dissertation, we explain this proof in detail, which can be distilled in three thematic cornerstones: demonstrating the existence of the threshold, the Main Lemma and the algorithmic construction of a 'cheap' cover. The Main Lemma implicitly relies tail bounds – in particular the Chernoff Bound. We study these in the second and third chapters, along with the notion of thresholds. We illustrate the strength of Kahn-Kalai conjecture with the problem of finding the threshold for the appearance of the Square of a Hamilton Cycle. We also introduce Park and Pham's reformulation of conjecture to better set-up the Main Lemma and carefully show that the reformulated conjecture implies the original Kahn-Kalai conjecture. Finally, we explain in detail the proof of the Main Lemma, and the randomized algorithm.

Chapter 2

Probability Background

In this chapter, we set up the discrete probability space and introduce the necessary concepts and results. We prove the Markov and Chernoff Inequalities, deriving a simplified version of the latter for subsequent implementation in the proof of the Main Lemma.

2.1 Discrete Probability Spaces

Proving results such as the existence of a Hamiltonian cycle in a Random Graph are often NP-Hard [Upfal]. Informally, this means that solving these problems is as computationally complex as solving NP problems, which require deterministic polynomial time to prove. An alternative approach is to consider the set of all edges as a discrete probability space. This opens the door to leveraging a set of proofs often pinned the 'Probabilistic Method'. The following definitions and results are borrowed from Mitzenmacher and Upfal (2005) and help build a formal framework to analyse random experiments.

Definition 1. (Discrete Probability Space) A *discrete probability space* is a triple $(\Omega, \mathcal{A}, \mathbb{P})$ consisting of the *sample space* Ω, which is the set of all possible outcomes of the random experiment, an *algebra* $\mathcal{A}$ representing the allowable events, which is family of subset of Ω containing the empty set and closed under complement and finite union, and a probability function $\mathbb{P} : \mathcal{A} \to [0, 1]$ satisfying Definition 2. An element of $\mathcal{A}$ is an event. An element in Ω is called a *simple* event.

The Kahn-Kalai conjecture is only concerned with discrete probability spaces where the sample space Ω is finite or countably infinite. The probability function will be uniquely defined by the probabilities of simple events (Mitzenmacher and Upfal (2005).

Definition 2. (Probability Function) A *Probability Function is any function* $\mathbb{P} : \mathcal{A} \to [0, 1]$ *such that for any event* $A \in \mathcal{A}$, $0 \leq \mathbb{P}(A) \leq 1$; $\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1$; *and given a (finite) disjoint sequence of* N *events* $A_1, A_2, \ldots$, *with* $A_i \in \mathcal{A}$ *for all* i,

$$\mathbb{P}(\bigcup_{i=1}^{N} A_i) = \sum_{i=1}^{N} \mathbb{P}(A_i) = \sum_{i=1}^{N} \sum_{\omega \in A_i} \mathbb{P}(\omega)$$

To provide some intuition for the following definition, we introduce a canonical example of a discrete probability space.

Example 3. Consider constructing a Random Graph from a vertex set $[n] := \{1, 2, \ldots n\}$ by selecting each edge uniformly $v_i v_j$ with $v_i, v_j \in [n]$ with some probability p. We associate to each edge the sample space $\Omega_{i,j} := \{A_1, A_2\}$ with A_1 the event where the edge is drawn and A_2 the alternative event where it is not drawn. We take the indicator variable $X : \Omega_{i,j} \to \{0, 1\}$. A Random Graph can be modelled by considering the random variables associated with each edge.

This requires the formal definition of a product space and product measure, which we borrow from Klenke (2020).

Definition 4. (Product Space, Product Measure) Let $\Omega_1, \Omega_2, \ldots \Omega_n$ be a series of spaces with associated Probability functions $\mathbb{P}_i : \Omega_i \to [0, 1]$. The *Product Space* $\Omega := \bigtimes_{i=1}^{n} \Omega_i$ *with probability function* $\mathbb{P} : \Omega \to [0, 1] := \prod_{i=1}^{n} \mathbb{P}_i$ *is a discrete probability space. For events* A_i *in* Ω_i, *we have the* product event $A := \bigtimes_{i=1}^{n} A_i$ *and* $\mathbb{P}(A) = \prod_{i=1}^{n} \mathbb{P}_i(A_i)$.

Following this principle, we also stipulate that the probability of drawing a given edge is not conditioned on that of drawing a different edge.

Definition 5. (Independence of Events) Let I be an arbitrary index set and let $(A_i)_{i \in I}$ be an arbitrary family of events in $\mathcal{A}$. The family $(A_i)_{i \in I}$ is independent if for any $J \subseteq I$,

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j)$$

When evaluating the outcome of a random process one often aims to associate some value to the event. We now recall standard concepts and results which are likely familiar to the reader, and refer to Mitzenmacher and Upfal (2005) and Klenke (2020) for more elaborate discussion.

Definition 6. (Random Variable) A *Random Variable* X is a measurable map $X : \Omega \to \mathbb{R}$. A *discrete random variable* is a Random variable which takes finitely or countably infinite number of values. If $X : \Omega \to \{0, 1\}$ we say it is an indicator variable.

We can now reformulate the event A: '$X = r$' as $A = \{\omega \in \Omega : X(w) = r, r \in \mathbb{R}\}$ and define the expected value of the random variable.

Definition 7. (Expectation)The *expectation* of a random variable X is

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega).$$

If X is an indicator variable and its expectation is $\mathbb{E}[X] = \sum_{\omega \in \Omega : X(\omega) = 1} \mathbb{P}(\omega)$.

A crucially useful property of the expectation is its linearity, which states that the expectation of a union of independent events can be computed by summing the respective expectations of individual events. More generally, we will use the independence of random variables.

Proposition 8. (Linearity of Expectation) Let $\{X_i\}_{i=1}^N$ be a series of N random variables on Ω and $\{\lambda_i\}_{i=1}^N \subseteq \mathbb{R}$. For an arbitrary, finite index set I, $X = \sum_{i\in I} \lambda_i X_i$ is also a random variable and

$$\mathbb{E}[X] = \sum_{i\in I} \lambda_i \mathbb{E}[X_i] = \sum_{i\in I} \sum_{\omega\in\Omega} \lambda_i X_i(x)\mu(\omega).$$

If X_i are indicator variables, $\mathbb{E}[X] = \sum_{i\in I} \lambda_i \sum_{\omega\in\Omega : X_i(\omega)=1} \mathbb{P}(\omega)$.

Indicator variables and their expectation are at the core of the probabilistic method, a set of non-constructive and countless Diestel (2017) proofs used to demonstrate the existence of objects which possess a property.

2.2 Tail Bounds

The probabilistic method is useful to demonstrate that a set possesses some required property. It stems from the following idea: given an indicator variable on a countable set X, we demonstrate that there exists some sample space of events with strictly positive probability that a random event will have the required property. For instance, if a random graph on n vertices contains a triangle with non-zero probability, then there must be at least one graph on n vertices containing a triangle. We formalize these principles below.

Lemma 9. *Given a sample space Ω and an indicator variable Y, if $\mathbb{E}[Y] > \alpha$, for some $\alpha > 0$, then there exists $\omega \in \Omega$ such that $Y(w) \geq \alpha$.*

Proof. Otherwise, the expectation of Y is upper bounded by α as follows.

$$\mathbb{E}[Y] = \sum_{\omega\in\Omega} Y(\omega)\mathbb{P}(\omega) < \sum_{\omega\in\Omega} \alpha\mathbb{P}(\omega) = \alpha.$$

□

Two omnipresent proof methods to demonstrate that some subset has a required property a.s. are the 'First' and 'Second' Moment methods; we illustrate these in the next chapter. They rely on tail bounds, which limit the deviation of a random variable Y_n its sample average $E[Y_n]$ from its expectation $E[Y]$. The following results are taken from Klenke (2020).

Lemma 10. *(Markov's Inequality) Let $Y \geq 0$ be a random variable on a discrete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\alpha > 0$. Then $\mathbb{P}(Y \geq \alpha) \leq \frac{\mathbb{E}(Y)}{\alpha}$. In particular, $\mathbb{P}(Y \geq 1) \leq \mathbb{E}(Y)$*

Proof. $\mathbb{E}(Y) \geq \mathbb{E}(Y \cdot \mathbb{1}_{Y\geq\alpha}) \geq \mathbb{E}(\alpha \cdot \mathbb{1}_{Y\geq\alpha}) = \alpha \cdot \mathbb{P}(\{\, \omega \in \Omega : Y(\omega) \geq \alpha \,\})$ □

The 'First Moment method' is a straightforward application of Markov's Inequality. The 'Second Moment method' is an application of Chebyshev's Inequality.

Lemma 11. *(Chebyshev's Inequality) Let Y be a random variable on a discrete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{E}(Y) < \infty$ and $Var(Y) < \infty$. Then for $k > 0$, $\mathbb{P}(|Y - \mathbb{E}(Y)| \geq k) \leq \frac{Var(Y)}{k^2}$.*

Proof. $\mathbb{P}\big[|Y - \mathbb{E}(Y)| \geq k\big] = \mathbb{P}\big[(Y - \mathbb{E}(Y))^2 \geq k^2\big] \overset{10}{\leq} \frac{\mathbb{E}[Y-\mathbb{E}(Y)]^2}{k^2} = \frac{Var(Y)}{k^2}$ □

This concentration bound is central to the 'Second Moment method', which as opposed to the 'First Moment method' can help provide a better bounds than Markov's Inequality.

Corollary 12. *(Second Moment Method) If Y is a non-negative, integer-valued random variable on the sample space Ω, then with $k = \mathbb{E}(Y)(> 0)$,*

$$\mathbb{P}(Y = 0) \leq \frac{Var(Y)}{(\mathbb{E}Y)^2} = \frac{\mathbb{E}(Y^2)}{(\mathbb{E}Y)^2} - 1$$

Proof. Set $k = \mathbb{E}Y$ in the Chebyshev Inequality. Then

$$\mathbb{P}(Y = 0) \leq \mathbb{P}(|\mathbb{E}Y - Y| \geq \mathbb{E}Y) = \mathbb{P}(|Y - \mathbb{E}Y| \geq \mathbb{E}Y) \leq \frac{Var(Y)}{(\mathbb{E}Y)^2}$$

□

However, the above-mentioned two bounds are often not 'strong' enough:we will require exponentially decreasing bounds on these probabilities. We take the following results from Frieze and Karoński (2015).

Lemma 13. *(Chernoff/Hoeffding Inequality)[CMU] Let $S_n = Y_1 + Y_2 + ... Y_n$ be random variables on $(\Omega, \mathcal{F}, \mu)$ a probability space where (i) $0 \leq Y_i \leq 1$ where $\mathbb{E}(Y_i) = \mu_i$, and (ii) Y_i's are independent. Let $\mu = \mu_1 + \mu_2 + ... + \mu_n$.*
Let $\phi(x) = (1 + x)\log(1 + x) - x$ if $x \geq -1$ and ∞ otherwise. Then for $\lambda \geq 0$,

$$\mathbb{P}(S_n \geq \mu + k) \leq e^{-\mu\phi(k/\mu)}$$

and

$$\mathbb{P}(S_n \leq \mu - k) \leq e^{-\mu\phi(-k/\mu)}$$

The subsequent bound is easier to apply in practice.

Corollary 14. *Putting $k = \epsilon\mu$, with $0 < \epsilon < 1$, we easily obtain the following convenient bounds.*

$$\mathbb{P}(S_n \geq (1 + \epsilon)\mu) \leq \left(\frac{e^{\epsilon}}{(1 + \epsilon)^{1+\epsilon}}\right)^{\mu} \leq \exp\left\{-\frac{\mu\epsilon^2}{3}\right\}$$

while

$$\mathbb{P}(S_n \leq (1 - \epsilon)\mu) \leq \exp\left\{-\frac{\mu\epsilon^2}{3}\right\}$$

This concentration bound is central to Park and Pham's reformulation of the original Kahn-Kalai conjecture.

Chapter 3

The Kahn-Kalai Conjecture

In this chapter, we introduce the notion of a threshold for a non-decreasing property and the conjecture enunciated by Kahn and Kalai[2016]. We then present and carefully justify the reformulated conjecture. Throughout this chapter we assume the following framework: consider a finite set X with cardinality n, and the sequence of Bernoulli trials in which each element of X is drawn with probability p. We take the product space $\Omega := \bigtimes_{i=1}^{n} \Omega_i$ satisfying Definition 4. The event $A \in \Omega$ 'for all $j \in J \subseteq X$, the j^{th} element is drawn' has probability $\mathbb{P}(A) = p^{|J|}(1-p)^{|X \setminus J|}$.

3.1 Threshold for Monotone Properties

To motivate the Kahn-Kalai conjecture, we first study the notion of thresholds for monotone properties. We first introduce Erdős–Rényi thresholds before considering a more stringent definition. In what follows, $\mathcal{F} \subseteq 2^X$ will always be a non-trivial increasing property.

Definition 15. An *increasing property* $\mathcal{F} \subseteq 2^X$ is a family of sets such that if $B \supseteq A \in \mathcal{F}$, then $B \in \mathcal{F}$. A property is *non-trivial* if $\mathcal{F} \neq \emptyset$.

Consider a uniformly drawn subset X_p of X, where each element is taken with probability p. Denote by $X_p \in \mathcal{F}$ the event 'there exists $A \in \mathcal{F}$ such that $X_p \subseteq A$' .

Lemma 16. *The event $X_p \in \mathcal{F}$ happens with probability*

$$\mathbb{P}(X_p \in \mathcal{F}) = \sum_{A \in \mathcal{F}} \mathbb{P}(A) = \sum_{A \in \mathcal{F}} p^{|A|}(1-p)^{|X \setminus A|}.$$

One striking aspect of random sets is the abrupt nature of appearance of monotone properties as p increases. In particular, denoting $\frac{p(n)}{q(n)} = o(1)$ by $q \gg p$, there always exists a function $p^*(n)$, such that $p \gg p^*$ ensures $X_p \in \mathcal{F}$ a.s. and, conversely, $p^* \gg p$ implies $\mathbb{P}(X_p \in \mathcal{F}) = 0$ a.s. . Such a function is called an Erdős–Rényi threshold.

Definition 17. An *Erdős–Rényi threshold* $p^*(n)$ for $\mathcal{F} \subseteq 2^X$, $|X| = n$, is a function $p^*(n)$ such that

$$\lim_{n\to\infty} \mathbb{P}(X_p \in \mathcal{F}) = \begin{cases} 0, & \text{if } p^*(n) \gg p(n) \\ 1, & \text{if } p(n) \gg p^*(n) \end{cases}$$

Erdős–Rényi threshold functions are not unique as any function $p(n)$ which differs from $p^*(n)$ by a constant factor is also a threshold. We will illustrate this definition with an example borrowed from Skerman (2008).

Example 18. Consider the graph $G_{n,p}$ where each edge ω is drawn uniformly from the set of edges X with probability p. A *cycle* is a walk without repeat vertices. Let $\mathcal{F}$ be the property that $G_{n,p}$ contains a cycle of any length. It is increasing and non-trivial, and we aim to find the associated Erdős–Rényi threshold.

Define Y_k to be the number of k-cycles cycles in the obtained graph, with $3 \leq k \leq n$. Each cycle is a fixed set of k edges C_k in $G_{n,p}$. Denote $C_k \in \mathcal{F}$ the event 'the random graph $G_{n,p}$ contains a k-cycle on the set of edges C_k'. We count over all possible sets of k edges $Y_k = \sum_{C_k} \mathbb{1}_{C_k \in \mathcal{F}}$. Noting the minimum requirement of 3 vertices to contain a cycle, we count the total number of cycles, Y_n by marginalizing over k.

$$Y_n = \sum_{k=3}^{n} Y_k = \sum_{k=3}^{n} \sum_{C_k} \mathbb{1}_{C_k \in \mathcal{F}}$$

Denote $(n)_k := n(n-1)...(n-k+1)$, then there are $(n)_k$ sequences $S = v_0, ...v_{k-1}$ of vertices in the labelled set $[n]$. Each cycle is identified by $2k$ such sequences as the following two functions yield the same cycle up to rotation and orientation.

$$f_1 : S \longrightarrow S \text{ defined by } f_1(v_i) = v_{i+x(\text{mod } k)}, \text{ and } f_2 : S \longrightarrow S \text{ defined by } f_2(v_i) = v_{k-1-i}.$$

Hence there are $\frac{(n)_k}{2k} = \binom{n}{k}\frac{(k-1)!}{2}$ such cycles and for a given $G_{n,p}$, $\mathbb{1}_{C_k \in \mathcal{F}}$ is Bernoulli with parameter p^k. By linearity of expectation, and noting $\binom{n}{k}(k-1)! < \binom{n}{k}k! < n^k$,

$$\mathbb{E}(Y_n) = \sum_{k=3}^{n} \mathbb{E}(Y_k) = \sum_{k=3}^{n} \sum_{C_k} \mathbb{P}(C_k)\mathbb{1}_{C_k \in \mathcal{F}} = \sum_{k=3}^{n} \frac{(n)_k}{2k} p^k < (np)^3 \sum_{k=0}^{n-3} (np)^k = \frac{(np)^n - (np)^3}{np - 1}$$

Using Markov's inequality, we can show that $G_{n,p} \notin \mathcal{F}$ if p is such that $\lim_{n\to\infty} np = 0$.

$$\lim_{n\to\infty} \mathbb{P}(Y_n \geq 1) \leq \lim_{n\to\infty} \mathbb{E}(Y_n) \leq \lim_{n\to\infty} \frac{(np)^n - (np)^3}{np - 1}$$

This implies the following lower bound on the probability of obtaining no cycle in $G_{n,p}$.

$$1 - \lim_{n\to\infty} \mathbb{P}(Y_n \geq 1) \geq 1 - \lim_{n\to\infty} \mathbb{E}(Y_n) \geq 1 - \lim_{n\to\infty} \frac{(np)^n - (np)^3}{np - 1}$$

Hence if p is such that $\lim_{n\to\infty} \frac{(np)^n}{1-np} = 0$, then $\lim_{n\to\infty} \mathbb{P}(G_{n,p} \in \mathcal{F}) = 0$ – which is the case if $p \ll 1/n$. This is referred to as the 'First Moment method'. Conversely, we would like to show that for $p \gg 1/n$, then $G_{n,p}$ contains a cycle almost surely. The Markov Inequality is not immediately useful: we will instead use Chebyshev's inequality.

Note that a sufficient condition for $G_{n,p} \in \mathcal{F}$ is that $G_{n,p}$ contains n edges. As the random graph $G_{n,p}$ is an $\binom{n}{2}$-dimensional random variable, where the entries are independent Bernoulli random variables $E_i \in \{0, 1\}$ with parameter p, the number of edges $Z_n = \sum_{i=1}^{n} E_i$ is distributed as $bin(\binom{n}{2}, p)$. As we would like to show that for $p \gg 1/n$ $G_{n,p}$ contains a cycle almost surely, it is sufficient to show this holds for $p = \frac{10}{n}$ due to monotonicity. Moreover, we note that $\mathbb{E}(Z_n) = \binom{n}{2}\frac{10}{n} = \frac{10}{n}\left(1 - \frac{1}{n}\right)$, $Var(Z_n) = \binom{n}{2}\frac{10}{n}$.

In fact, to show $Z_n > n$ a.s., it is sufficient to show that $Z_n > \mathbb{E}(Z_n) - \frac{n}{10}$ a.s. . Indeed if $Z_n > \frac{49}{10}n - 5$, then it can easily be shown by induction that with $n \geq 3$, we have $Z_n > n$. Using Chebyshev's inequality with $k = \frac{10}{n}$ we obtain a bound on the distance between Z_n and its expectation.

$$\mu_p(|\mathbb{E}(Z_n) - Z_n| \geq \frac{n}{10}) \leq \frac{Var(Z_n)}{(n/10)^2} = \frac{10^3}{2n} \cdot (1 - \frac{1}{n}) \cdot (1 - \frac{10}{n}) \to 0$$

Hence as $p = 10/n$, $G_{n,p} \in \mathcal{F}$ a.s. as it contains n edges a.s.. Due to monotonicity, this is also holds for $p \gg 1/n$. We conclude that $p^*(n) = \frac{1}{n}$ is an Erdős–Rényi threshold for $\mathcal{F}$. This is often referred to as the 'Second Moment method'.

It can be shown that every monotone, non-trivial property admits a threshold in the Erdős-Rényi sense. In 1987, Bollobás and Thomason prove an even stronger result Bollobás and Thomason (1987).

3.2 The Bollobàs-Thomason Theorem

In this section we state the Bollobás-Thomason theorem and outline the proof. From now on we consider a slightly finer definition of threshold, as do the authors of Park and Pham (2022).

Definition 19. A function $p^* = p^*(n)$ is a *threshold* for a monotone increasing property $\mathcal{F} \subseteq 2^X$ is the unique p^* such that $\mathbb{P}(X_{p^*} \in \mathcal{F}) = 1/2$ a.s. .

To discuss the existence of such thresholds, the following 'coupling technique' will be useful. Intuitively, it is plausible that if $p_1 < p_2$, then it is less likely that $X_{p_1} \in \mathcal{F}$ than $X_{p_2} \in \mathcal{F}$. The next lemma formally asserts this.

Proposition 20. (Coupling Technique) Consider the sequence of random subsets of X $X^1_{p_1}, X^2_{p_2}, \ldots X^k_{p_k}$, each drawn independently to iteratively construct $X_p := \bigcup_{i=1}^k X^i_{p_i}$. Then $p = 1 - \prod_{i=1}^k (1 - p_i)$. Moreover, for $0 < p_1 < p_2 < 1$, then $\mathbb{P}(X_, p_1 \in \mathcal{F}) \leq \mathbb{P}(X_{n,p_1} \in \mathcal{F})$.

Proof. Let $\omega_i \in X, 1 \leq i \leq n$ be any fixed element of Xassociated with sample space Ω_i with the indicator random variable $Z^i_j : \Omega_j \to \{0, 1\}$ which takes value 1 if ω_j is drawn in the realization of $X^i_{p_i}$ and 0 otherwise.

$$\mathbb{P}(\omega_j \notin X_p) = \mathbb{P}\left(\bigcap_{i=1}^k (Z^i_j)^{-1}(\{0\})\right) = \prod_{j=1}^n \mu_i\left((Z^i_j)^{-1}(\{0\})\right) = \prod_{j=1}^k (1 - p_j)$$

In particular, this says that if $0 < p_1 < p_2 < 1$, then $\mathbb{P}(X_, p_1 \in \mathcal{F}) \leq \mathbb{P}(X_{n,p_1} \in \mathcal{F})$. Indeed, we can choose $0 < q < 1$ such that $p_2 = 1 - (1 - p_1)(1 - q)$, and obtain from the coupling of independently drawn X_{p_1}, X_q .

$$\begin{aligned}
\mathbb{P}(X_{p_2} := X_{p_1} \cup X_q \notin \mathcal{F}) &= \mathbb{P}(X_{1-(1-p_1)(1-q)} \notin \mathcal{F}) \text{ , by 20} \\
&= \mathbb{P}((X_{n,p_1} \notin \mathcal{F}) \cap (X_{n,q} \notin \mathcal{F})) \text{ , by definition} \\
&= \mathbb{P}((X_{n,p_1} \notin \mathcal{F}))\mathbb{P}((X_{n,q} \notin \mathcal{F})) \text{ , by indendence} \\
&< \mathbb{P}((X_{n,p_1} \notin \mathcal{F}))
\end{aligned}$$

We can conclude $\mathbb{P}(X_{p_2} \in \mathcal{F}) > \mathbb{P}(X_{p_1} \in \mathcal{F})$.

□

The following theorem states every non-trivial monotone property $\mathcal{F}$ admits a threshold function p^*. Furthermore, this threshold is in some sense 'abrupt': for some scalar multiple kp^* below p^* then w.h.p. $X_p \notin \mathcal{F}$. Symmetrically for some Kp^*, above p^* then w.h.p. $X_p \in \mathcal{F}$ (Bollobás and Thomason (1987)).

Theorem 21. *Bollobás-Thomason Let $\mathcal{F}$ be non-trivial increasing. For every $0 < \epsilon \leq 1/2$, there exists a unique p^* and some constants $K_\epsilon^1, K_\epsilon^2$ such that*

$$\mathbb{P}(X_{p^*} \in \mathcal{F}) = 1/2, \mathbb{P}(X_{k_\epsilon^2 p^*} \notin \mathcal{F}) \leq \epsilon \text{ and } \mathbb{P}(X_{K_\epsilon^2 p^*} \in \mathcal{F}) \geq 1 - \epsilon.$$

Proof. We first show existence and uniqueness. For $p(n) \in [0,1]$, the function givnen by $\mathbb{P}(X_p \in \mathcal{F}) = \sum_{A \in \mathcal{F}} p^{|A|}(1-p)^{|X \setminus A|}$ is a polynomial in p and hence continuous. Non-triviality implies $\mathbb{P}(X_0 \in \mathcal{F}) = 0$ and $\mathbb{P}(X_1 \in \mathcal{F}) = 1$. By the Intermediate Value Theorem, the existence of p^* such that $\mathbb{P}(X_{p^*} \in \mathcal{F}) = 1/2$. is garanteed. Moreover, p^* is unique as $\mathbb{P}(X_p \in \mathcal{F})$ is strictly increasing in p. Indeed suppose $0 < p_1 < p_2 < 1$. then by the coupling technique, we have $\mathbb{P}(X_{n,p_1}) < \mathbb{P}(X_{n,p_2})$.

Now let ϵ be given. Chosoe $k_1(\epsilon) > -\log_2(\epsilon)$. Consider the sequence of

$X_{p^*}^1, X_{p^*}^2, \ldots X_{p^*}^{k_1(\epsilon)}$, each drawn independently. Let $X_{1-(1-p^*)^{k_1(\epsilon)}} := \bigcup_{i=1}^{k_1(\epsilon)} X_{p^*}$ be constructed iteratively. We note that $1-(1-p^*)^{k_1(\epsilon)} < p^* k_1(\epsilon)$ and hence by the coupling argument, $\mathbb{P}(X_{p^* k_1(\epsilon)} \notin \mathcal{F}) < \mathbb{P}(X_{1-(1-p)^{k_1(\epsilon)}} \notin \mathcal{F})$. Further, we observe that if $X_{1-(1-p^*)^{k_1(\epsilon)}} \notin \mathcal{F}$, then $\bigcup_{i=1}^{k_1(\epsilon)} X_{p^*} \notin \mathcal{F}$, which in turn implies that $\bigcap_{i=1}^{k_1(\epsilon)} X_{p^*} \in \mathcal{F}$. Together, these observations give us the following bound.

$$\begin{aligned} \mathbb{P}(X_{p^* k_1(\epsilon)} \notin \mathcal{F}) < \mathbb{P}(X_{1-(1-p^*)^{k_1(\epsilon)}} \notin \mathcal{F}) &\leq \mathbb{P}\left(\bigcap_{i=1}^{k_1(\epsilon)} X_{p^*} \in \mathcal{F}\right) \\ &= \mathbb{P}(X_{p^*} \in \mathcal{F})^{k_1(\epsilon)} = (1/2)^{k_1(\epsilon)} < \epsilon. \end{aligned}$$

To vindicate the second inequality, choose $k_2(\epsilon) > -\log_2(\epsilon) + 1$ and pick $q = 1-(1-p^* k_2(\epsilon)(1-p^*)^{1-k_2(\epsilon)}$ such that $k_2(\epsilon)p^* = 1-(1-p^*)^{k_2(\epsilon)-1}(1-q)$. This entails $k_2(\epsilon)p^* > 1-(1-p^*)^{k_2(\epsilon)-1}$, then again by the coupling argument we obtain the desired bound.

$$\begin{aligned} \mathbb{P}(X_{p^* k_2(\epsilon)} \notin \mathcal{F}) < \mathbb{P}(X_{1-(1-p^*)^{k_2(\epsilon)-1}} \notin \mathcal{F}) &\leq \mathbb{P}\left(\bigcap_{i=1}^{k_2(\epsilon)-1} X_{p^*} \in \mathcal{F}\right) \\ &= \mathbb{P}(X_{p^*} \in \mathcal{F})^{k_2(\epsilon)-1} = (1/2)^{k_2(\epsilon)-1} < \epsilon. \end{aligned}$$

□

3.3 The Kahn-Kalai Conjecture

While it ensures existence of a threshold function p^*, the Bollobás-Thomason Theorem does not provide information on the location of p^*. One intuitive starting point to find these threshold function is the 'expectation threshold'. We henceforth consider base 2 log functions, and illustrations are due to Rao (2023).

A key concept behind the expectation threshold is that of a *p-small cover* . To provide some intuition, consider Figure 3.1. X_p is still a uniformly random subset of X. The black dots represent the sets in $\mathcal{F}_{min} := \{ F \in \mathcal{F} : F \text{ minimal} \}$.

Definition 22. A *minimal* element G of a set $\mathcal{G}$ is such that $\forall G' \in \mathcal{G}$, then $G \subsetneq G'$.

The points contained in the 'closure' of the cones is $\mathcal{F} = \langle \mathcal{F}_{min} \rangle$. We are interested in the size of $\mathcal{F}_p := \{ F \in \mathcal{F} : F \subseteq X_p \}$. The idea is to approximate the set $\mathcal{F}_p$ with one of lower expected 'size'. Indeed, Figure 3.2 shows that $\mathcal{F}_p \subseteq \langle G \rangle$, but, crucially, the sets in G have less overlap.

Definition 23. (p-small, Cover) We say that $\mathcal{F}$ is *p-small* if there exists $\mathcal{G} \subseteq 2^X$ such that both of the following hold.

$$\mathcal{F} \overset{(1)}{\subseteq} \langle \mathcal{G} \rangle := \bigcup_{S \in \mathcal{G}} \left\{ T \subseteq 2^X : S \subseteq T \right\} \text{ and } \sum_{S \in \mathcal{G}} p^{|S|} \overset{(2)}{\leq} \frac{1}{2}$$

If (1) holds, we say $\mathcal{G}$ is a *cover* of $\mathcal{F}$.

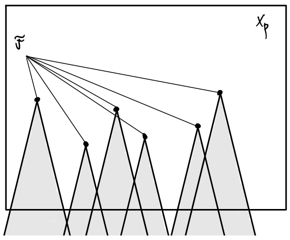

Figure 3.1: Increasing Property $\mathcal{F}$ and its Minimal Elements

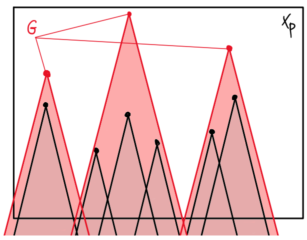

Figure 3.2: Minimal elements for a Cover $\mathcal{G}$ of $\mathcal{F}$

We recognize the second requirement in the definition of *p-small* to be the expected size of the set $\mathcal{G}_p := \{S \in \mathcal{G} : S \subseteq X_p\}$. Consider product space $\Omega := \bigtimes_{i=1}^{n} \Omega_i$ satisfying Definition 4. The event $A_i \in \Omega_i = \{ A_i, A_i^C \}$ is 'the element $j \in X$ is drawn in the realization of X_p, and the associated indicator variable is $Y : \Omega \to \{ 0, 1 \}^n$.

$$\mathbb{E}[|\{S \in \mathcal{G} : S \subseteq X_p\}|] \overset{(1)}{=} \sum_{i=1}^{|\mathcal{G}|} \mathbb{E}\left[\mathbb{1}_{S_i \subseteq X_p}\right] \overset{(2)}{=} \sum_{i=1}^{|\mathcal{G}|} \mathbb{E}\left[\prod_{j=1}^{|S_i|} \mathbb{1}_{S_i^j \in X_p}\right] \overset{(3)}{=} \sum_{i=1}^{|\mathcal{G}|} \prod_{j=1}^{|S_i|} \mathbb{E}[\mathbb{1}_{S_i^j \in X_p}]$$
$$\overset{(4)}{=} \sum_{i=1}^{|\mathcal{G}|} \prod_{j=1}^{|S_i|} \mathbb{P}\left(Y^{-1}(S_i^j)\right) \overset{(5)}{=} \sum_{i=1}^{|\mathcal{G}|} p^{|S_i|}$$

In the above (1) is due to linearity of expectation, (2) denotes S_i^j the j^th element of S_i, (3) uses the independence of Bernoulli trials, (4) follows from the fact that $\mathbb{1}_{S_i^j \in X_p}$ are simple functions, and (5) is a result of $\mathbb{P}\left(Y^{-1}(S_i^j)\right) = p$.

Definition 24. The *expectation threshold* $q(\mathcal{F})$ of $\mathcal{F}$ is the maximum p such that $\mathcal{F}$ is p-small.

Intuitively, $q(\mathcal{F})$ is the largest p such that the expected size of the cover G in figure 3.2 is no more than $1/2$. One immediately notices that the expectation threshold provides a trivial lower bound on the threshold, as shown below.

$$\mathbb{P}(X_p \in \mathcal{F}) \leq \mathbb{P}(X_p \in \langle G \rangle) \leq \sum_{S \in \mathcal{G}} p^{|S|},$$

For an arbitrary $p < q(\mathcal{F})$, $\mathbb{P}(X_p \in \mathcal{F}) < \frac{1}{2}$, preventing p from being the threshold p^*. Given the diversity and complexity of certain properties, finding a universal upper bound on p^* seems radical. In 2006, Kahn and Kalai Kahn and Kalai (2007) advanced the following conjecture. Let l_0 be the largest minimal element of $\mathcal{F}$, and let $l := \max\{l_0, 2\}$.

Theorem 25. *(The Kahn-Kalai Conjecture) For every finite set X and non-trivial increasing property $\mathcal{F} \subseteq 2^X$, there exists a universal constant K such that*

$$p^*(\mathcal{F}) \leq Kq(\mathcal{F}) \log l(\mathcal{F}).$$

This suggests that for any $\mathcal{F}$, the 'trivial' lower bound for $p^*(\mathcal{F})$ is always within a factor $O(log(l))$ of the truth. To emphasize the importance of this theorem, Park and Pham recall that Kahn and Kalai themselves stated, regarding: 'It would probably be more sensible to conjecture that it is not true; but it does not seem easy to disprove' Kahn and Kalai (2007).

3.4 Squares of Hamilton Cycles

The study of Hamilton Cycles and thresholds for their appearance has been central to the study of random graphs. A Hamilton Cycle is a graph C on $[n]$ which is a cycle containing all vertices in $[n]$ exactly once. It is a well known result, established by Pósa (1976) that the threshold for the appearance of a Hamilton Cycle is of order $\log(n)/n$. Fascinatingly, it is an isolated case of a class of graphs named 'Powers of Hamilton Cycles'. The following definition is taken from Montgomery (2018).

Definition 26. (Power of Hamilton Cycle) The k^{th} power of C, C^k is the cycle graph with vertex set $[n]$ and edge set $E(H) = uv : d_H(u,v) \leq k$, where $d_H(u,v) \leq k$ is the length of a shortest u, v-path in C if one exists.

Riordan (2000) demonstrated that for all $k \geq 3$, if $p = O(n^{-1/k})$, then $G_{n,p}$ contains a k-cycle C_n^k a.s. using an argument based on the 'Second Moment' method. Indeed, this corresponds to the the expected number of C_n^k in $G_{n,p}$, which is $\frac{1}{2}(n-1)!p^{kn}$. However, the case for $k = 2$ has remained open for a longer period of time. Kühn and Osthus (2012) conjectured that the threshold is $n^{1/2}$, and the best bounds before 2020 were of $(\log n)^3 n^{-1/2}$ by Fischer, Škorić, Steger, and Trujić (2022) and of $(\log n)^2 n^{-1/2}$ by Friedgut (2005).

However, the proof of the Kahn-Kalai conjecture allows us to easily find an upper-bound within a logarithmic factor of the true threshold for C_n^2. Indeed, it reduces this task to demonstrating that $\mathcal{H}_{C_{n^2}}$, the hypergraph consisting of all graphs on n vertices containing the square of a Hamilton cycle, is not q-small, where q is the expectation threshold. We first require a formal link between q-smallness and 'spread' which we borrow from Frankston, Kahn, Narayanan, and Park (2021).

Definition 27. (Spread Hypergraph, Spread Measure) Let $\mathcal{H}$ be an l-bounded hypergraph. Recall $\langle S \rangle = \{ T : S \subseteq T \subseteq X \}$. We say a probability measure μ on the edges of $\mathcal{H}$ is κ-spread if $\forall S \subseteq X$ we have

$$\mu(\{ A \in \mathcal{H} : S \subseteq A \}) = \mu(|\mathcal{H} \cap \langle S \rangle|) < \frac{1}{\kappa^{|S|}}.$$

In particular, we say that $\mathcal{H}$ is κ-spread if the uniform distribution κ-spread, i.e. $\forall S \subseteq X$ we have

$$\frac{|\mathcal{H} \cap \langle S \rangle|}{|\mathcal{H}|} \leq \frac{1}{\kappa^{|S|}},$$

where edges are counted with multiplicities on both sides.

We can now state the relationship between spread and 'smallness'.

Lemma 28. *Given an l-bounded hypergraph $\mathcal{H}$, if $\mathcal{H}$ is κ-spread then $\mathcal{H}$ is not κ^{-1}-small.*

Proof. Let $\mathcal{G}$ be a cover of $\mathcal{H}$, i.e. $\mathcal{H} \subseteq \langle \mathcal{G} \rangle$. As $\mathcal{H}$ is κ-spread, we have for all $G \in \mathcal{G}$,

$$\frac{|\mathcal{H} \cap \langle G \rangle|}{|\mathcal{H}|} = \sum_{\substack{A \in \mathcal{H} \\ G \subseteq A}} \mu(A) \leq \frac{1}{\kappa^{|G|}}.$$

Taking the sum over all elements G in the cover $\mathcal{G}$ we obtain the following bound.

$$\sum_{G \in \mathcal{G}} \frac{1}{\kappa^{|G|}} \geq \sum_{G \in \mathcal{G}} \sum_{\substack{A \in \mathcal{H} \\ G \subseteq A}} \mu(A) \geq 1,$$

where the last inequality follows from the fact that every set in $\mathcal{H}$ contains a set in $\mathcal{G}$. In particular, $\sum_{G \in \mathcal{G}} \frac{1}{\kappa^{|G|}} > 1/2$ so $\mathcal{G}$ is not κ^{-1} small. □

This concept allows us to use an equivalent formulation of Theorem 32.

Corollary 29. *Let $\mathcal{H}$ be an l-bounded, κ-spread hypergraph, for which the edges are subsets of X. There is an absolute constant $K > 0$ such that if $p \geq \frac{(K \log r)}{\kappa}$, then w.h.p. X_p contains an edge of $\mathcal{H}$. Here w.h.p. assumes that $l \to \infty$.*

It is left to show that the property '$G'_{n,p} \in \mathcal{H}_{C_{n^2}}$ is not a κ-spread hypergraph.

Proposition 30. $\mathcal{H}_{C_n^2}$ is κ-spread.

Proof. The proof is taken from Frieze and Karoński (2015). Given $\mathcal{H}_{C_n^2}$, we can obtain the following bound on the uniform distribution.

$$\frac{|\mathcal{H}_{C_n^2} \cap \langle S \rangle|}{|\mathcal{H}_{C_n^2}|} \leq \frac{(n-2-\lfloor |S|/2 \rfloor)!}{(n-1)!/2} \leq \left(\frac{e}{n-1}\right)^{\lfloor |S|/2 \rfloor + 1}$$

We therefore take $\kappa = e^{-1} n^{1/2}$ and immediately obtain that $\mathcal{H}_{C_n^2}$ is κ-spread. □

Proposition 30 shows that $\mathcal{H}_{C_n^2}$ is κ-spread and hence the only assumption of Corollary 29 is satisfied. Using Proposition 35, we can assert both of the following statements:

$\boxed{1}$ Let $\mathcal{H}_{C_n^2}^{min}$ be the set of minimal elements of $\mathcal{H}_{C_n^2}$. Theorem 32 implies there exists a universal constant L such that a uniformly random $(Lp \log l)|X|$ - element subset of X belongs to $\langle \mathcal{H}_{C_n^2}^{min} \rangle$ with probability $1 - o(1)$.

$\boxed{2}$ Noting that the largest element of $\mathcal{H}_{C_n^2}$ has size at most $l(\mathcal{H}_{C_n^2}) := n + \lfloor n/2 \rfloor = O(n)$, 25 implies there exists a universal constant K such that

$$n^{-1/2} \leq K n^{-1/2} \log l(\mathcal{H}_{C_n^2}) = O(n^{-1/2} \log n)$$

Park and Pham, through a more careful argument, show that the true threshold is in fact $n^{-1/2}$ in a paper which precedes their proof of the Kahn kalai Conjecture Kahn, Narayanan, and Park (2021).

Chapter 4

Proof of the Kahn-Kalai Conjecture

In the previous chapter, we underlined the importance of Theorem 25 by noting that Kahn and Kalai (2007) found a refutation of the conjecture more plausible than a vindication. We now present the key result of Park and Pham (2022); namely, the proof of Theorem **??** which implies Theorem 25, and follows significant advances in the field, in particular the proof Sunflower Conjecture by Alweiss et. al. Alweiss, Lovett, Wu, and Zhang (2021) and the proof the 'Fractional Kahn-Kalai Conjecture' Frankston, Kahn, Narayanan, and Park (2021).

4.1 Reformulation

In the this section we introduce Theorem 32, a reformulation of Theorem 25, and explain how the former implies the latter. The original conjecture considers properties, which are instances of hypergraphs.

Definition 31. (Hypergraph) A *hypergraph* on X is a collection $\mathcal{H}$ of subsets of X. A set $H \in \mathcal{H}$ is an *edge* of $\mathcal{H}$. We extend to hypergraphs the following notions: $\mathcal{H}$ is l-bounded if each edge has size at most l, and noting $\langle \mathcal{H} \rangle := \bigcup_{S \in \mathcal{H}} \{T \subseteq 2^X : S \subseteq T\}$, $\mathcal{H}$ is p-small if $\sum_{S \in \mathcal{H}} p^{|S|} \leq \frac{1}{2}$.

Below we state Park and Pham's reformulation of Theorem 25, which reduces the task of demonstrating that with $q = Lp \log l$ for some constant L, the random subset $X_q \in \langle \mathcal{F}_{min} \rangle$ with probability $1 - o_{l \to \infty}(1)$ to showing that the hypergraph $\mathcal{F}$ is not p-small. Importantly, showing $\sum_{S \in \mathcal{F}} p^{|S|} \leq \frac{1}{2}$ is generally a tractable problem.

Theorem 32. *(Reformulation of the Kahn-Kalai Conjecture) Let $l \geq 2$. There exists a universal constant L such that for any nonempty, l-bounded hypergraph $\mathcal{H}$ on X which is not p-small, a uniformly random $(Lp \log l)|X|$-element subset of X belongs to $\langle \mathcal{H} \rangle$ with probability $1 - o(1)$.*

In particular, we will obtain a quantitative lower bound of $log(l)^{-c}$ for $c > 0$ for the $o(1)$ term. proving the implication will require the following two lemmas. The first lemma is due to Frieze and Karoński (2015).

Lemma 33. *Let $X_p \subseteq X$ be as in the previous chapter, and X_m be a set chosen uniformly at random from all $m = p|X|$-sized subsets of X. Then $\mathbb{P}(X_p \in \mathcal{F}) > \mathbb{P}(X_m \in \mathcal{F} \big| |X_p| = m)$.*

Proof.

$$
\begin{aligned}
\mathbb{P}(X_p \in \mathcal{F}) &= \sum_{m=1}^{n} \mathbb{P}(X_p \in \mathcal{F} \big| |X_p| = m)\mathbb{P}(|X_p| = m) \text{ by the Law of Total Probability} \\
&= \sum_{m=1}^{n} \frac{\mathbb{P}((X_p \in \mathcal{F} \big| |X_p| = m) \wedge \mathbb{P}(|X_p| = m))}{\mathbb{P}(|X_p| = m)} \mathbb{P}(|X_p| = m) \\
&= \sum_{m=1}^{n} \sum_{\substack{F \in \mathcal{F} \\ |F| = m}} \frac{p^m (1-p)^{n-m}}{\binom{n}{m} p^m (1-p)^{n-m}} \mathbb{P}(|X_p| = m) \text{ by definition of } X_p \\
&= \sum_{m=1}^{n} \sum_{\substack{F \in \mathcal{F} \\ |F| = m}} (\binom{n}{m})^{-1} \mathbb{P}(|X_p| = m) \\
&= \sum_{m=1}^{n} \mathbb{P}(X_m \in \mathcal{F}) \mathbb{P}(|X_p| = m) \\
&\geq \mathbb{P}(X_m \in \mathcal{F}) \mathbb{P}(|X_p| = m)
\end{aligned}
$$

□

While Lemma 33 allows for comparison of $\mathbb{P}(X_p \in \mathcal{F})$ and $\mathbb{P}(X_m \in \mathcal{F})$, the following lemma is concerned only with the X_m model.

Lemma 34. *Let X be a finite set and X_m be a set chosen uniformly at random from all m-sized subsets of X. Then for $m' > m$, $\mathbb{P}(X_{m'} \in \mathcal{F}) \geq \mathbb{P}(X_m \in \mathcal{F})$.*

Proof. Denote $\mathcal{F}_m := \{ X \in \mathcal{F}_m : |X| = m \}$. We first note that for an arbitrary m, $\mathbb{P}(X_m \in \mathcal{F}) = \sum_{\substack{X' \subseteq [n] \\ |X'| = m \\ X' \in \mathcal{F}}} (\binom{n}{m})^{-1} = \frac{|\mathcal{F}_m|}{\binom{n}{m}}$. In addition, $|\mathcal{F}_{m'}| \geq |\mathcal{F}_m| \frac{\binom{n-m}{m'-m}}{\binom{m'}{m'-m}}$. Rearranging and multiplying both sides by $\frac{n!(m'-m)!}{m'!(n-m)!} = \binom{n}{m}(\binom{m'}{m'-m})^{-1}$ yields $|\mathcal{F}_{m'}|\binom{n}{m} \geq |\mathcal{F}_m|\binom{n}{m'}$. □

Using these two lemmas we can now show that Theorem 32 implies Theorem 25.

Proposition 35. Theorem 32 implies Theorem 25.

Proof. Let $\mathcal{F}$ be as in theorem 25, with $p^*(\mathcal{F})$ the true threshold, and $q(\mathcal{F})$ the expectation threshold. Assume $\tilde{q} > q(\mathcal{F})$. We will show that with $p = K\tilde{q} \log l(\mathcal{F})$ for some constant K to be determined, $\mathbb{P}(X_p \in \mathcal{F}) > \frac{1}{2}$ and hence $p^* < p$.

Define the hypergraph $\mathcal{H}$ to be the set of minimal elements of $\mathcal{F}$, such that $\langle \mathcal{H} \rangle = \mathcal{F}$. Note that $\mathcal{H}$ is l-bounded, and as $\tilde{q} > q(\mathcal{F})$, we have $\sum_{S \in \mathcal{H}} p^{|S|} > \frac{1}{2}$ and hence $\mathcal{F}$ is not $\tilde{q}$-small.

From theorem 32, one can choose a large enough L such that for $X_m \subseteq X$, an m-sized uniformly random subset of X with $m = (L\tilde{q} \log l)|X|$, will not belong to $\mathcal{H}$ with exceptional probability $o_{l\to\infty}(1)$. In particular, with $L = Cl(\mathcal{F})$ for some constant C large enough, we have $\mathbb{P}(X_m \notin \langle \mathcal{H} \rangle) < \frac{1}{4}$.

Now, let $p' = 2L\tilde{q} log l$, with K sufficiently large such that $K\tilde{q} \log l(\mathcal{F}) = p > p'$. If $p' > 1$, $\mathbb{P}(X'_p \in \langle \mathcal{F} \rangle) = 1$ and Theorem 25 holds trivially. Without loss of generality, suppose $p' < 1$. Letting X_m be a uniformly random, m-sized subset of X,

$$
\begin{aligned}
\mathbb{P}(X_{p'} \in \mathcal{F}) &\geq \sum_{m' \geqslant m} \mathbb{P}(X_{p'} \in \mathcal{F} \big| |X_{p'}| = m')\mathbb{P}(|X_{p'}| = m') \text{ by law of Total probability} \\
&\geq \sum_{m' \geqslant m} \mathbb{P}(X_{m'} \in \mathcal{F})\mathbb{P}(|X_{p'}| = m') \text{ by Lemma 33} \\
&\geq \mathbb{P}(X_m \in \mathcal{F}) \sum_{m' \geqslant m} \mathbb{P}(|X_{p'}| = m') \text{ by Lemma 34} \\
&= \mathbb{P}(X_m \in \mathcal{F})\mathbb{P}(|X_{p'}| \geqslant m) \text{ by independence of Bernoulli trials} \\
&\geq (1 - 1/4)\mathbb{P}(|X_{p'}| \geq m) \text{ by the result obtained above} \\
&\geq (3/4)\mathbb{P}(|X_{p'}| \geq m).
\end{aligned}
$$

Using the concentration bound obtained from Chernoff bounds in Corollary 14, we take $\mu = \mathbb{E}(|X_{p'}|) = p'|X| = 2(L\tilde{q} log l)|X| = 2m$, and $\delta = 1/2$, and we note that $\{\{x\} : x \in X\}$ covers $\mathcal{F}$ with $p' > p = K\tilde{q} \log l(\mathcal{F}) > \tilde{q}$, which entails that

$$|X|p' \geqslant |X|\tilde{q} = \mathbb{E}[|X_{\tilde{q}}|] = \sum_{i=1}^{|X|} \tilde{q} > 1/2$$

where we use the fact that $\mathcal{F}$ is not $\tilde{q}$-small. In particular, this implies m is somewhat large as $m = (L\tilde{q} \log l)|X| > (L \log l)/2$. We apply Corollary 14 to bound below $\mathbb{P}(|X_{p'}| \geq m)$.

$$\mathbb{P}(|X_{n,p'}| \leqslant 1/2 \cdot \mathbb{E}[|X_{n,p'}|]) \leqslant e^{-(2m)\delta^2/2} = e^{-m/4}$$

Rearranging, we obtain

$$\mathbb{P}(|X_{p'}| \geqslant 1/2 \cdot 2m) \geqslant 1 - e^{-m/4} > 2/3$$

And we can now conclude

$$\mathbb{P}(X_{p'} \in \mathcal{F}) \geq (3/4)\mathbb{P}(|X_{p'}| \geq m) \geq 1/2.$$

□

Theorem 32 implies that constructing a cover $\mathcal{U}$ for $\mathcal{H}$ with is not 'cheap', i.e.

$$\mathbb{E}[|\{S \in \mathcal{U} : S \subseteq X_p\}|] \leq 1/2 \text{ w.h.p.},$$

would suffice in order to show that a uniformly random $(Lp \log l)|X|$-element subset of X belongs to $\langle \mathcal{H} \rangle$ with probability $1 - o_{l\to\infty}(1)$.

4.2 Constructing a Cover

In this section we present the first component of the proof of Theorem 32. The overarching strategy is to iteratively construct two sets of elements: $\bigcup_i \mathcal{U}_i$ and $W = \bigcup_i W_i$ through a randomized iterative process. This algorithmic construction has two outputs: either $\bigcup_i \mathcal{U}_i$ is a p-small cover of $\mathcal{H}$ or W satisfied the conclusion of Theorem 32: it is a uniformly random $(Lp\log l)|X|$-element subset of X belongs to $\langle\mathcal{H}\rangle$ w.h.p..

To motivate the method, consider sampling uniformly random and disjoint W_i from $\binom{X}{w_i}$, the set of $w_i = L_i p|X|$-sized subsets of X at each step i of the process. Figure 4.1 depicts an ideal scenario: for some $S_1 \in \mathcal{F}$, $|S_1 \cap W_1| = 0.1 \cdot |S_1|$, $|S_1 \cap W_2| = 0.1 \cdot |S_1 \setminus W_1|$, ... and $|S_1 \cap W_2| = 0.1 \cdot |S_1 \setminus \cup_{j\le i} W_j|$. The number of steps k until S is fully covered is $O(\log_{0.9} l)$ as we require $|S_1 \setminus \cup_{j\le k} W_j| = 0.9^k \cdot |S_1| \le 1$ and assuming S is l-bounded, one obtains $k \le \log_{0.9}(1/l)$. At the end, the process yields $W \subseteq S \in \mathcal{H}$ where $W \in \binom{X}{m}$, with $m = O(Lpn\log l)$. This motivates the appearance of the $\log l$ in Theorem 32. Evidently, the typical behavior of $|S_1 \cap W_i|$ will be an Lp fraction of $|S_1|$, which is much smaller than 10% of $|S_1|$.

The following concept of 'minimum fragment' is a technical cornerstone in the construction of the p-small cover.

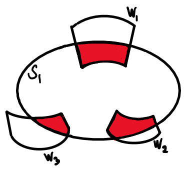

Figure 4.1: Idealized Piece-Wise Covering of $S \in \mathcal{F}$

Definition 36. (Minimum Fragment). Say we have $W \in \binom{X}{w}$, and some hypergraph $\mathcal{H} \subseteq 2^X$. Given any element S of $\mathcal{H}$, then $T(S,W)$ is a *minimum* (S,W)*-fragment if* $T(S,W) = S' \setminus W$ *for some* $S' \in \mathcal{H}$ *such that* $S' \subseteq S \cup W$ *and* $|S' \setminus W|$ *is smallest. We denote the size of a* minimum fragment$|T(S,W)|$ *by* $t(S,W)$.

Below we state a property of minimal fragments which provides some additional intuition and which will be crucial in the proof that follows.

Proposition 37. Let $T(S,W)$ and $S' \in \mathcal{H}$ be as in Definition 36. Then $T(S,W) \subseteq S'$.

Proof. S' is contained in $T \cup W \subseteq S \cup W$. If $T(S,W) \subsetneq S'$, then $S' \subseteq S \cup W$ implies that $|S' \setminus T| < |T|$. But this contradicts the minimality of T. □

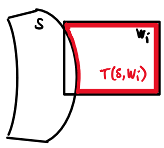

Figure 4.2: Typical Minimum $T(S,W_i)$ Fragment

Figure 4.2 depicts a typical minimal fragment, while Figure 4.3 shows that for disjoint W and S, one of two cases is possible. For all $S' \in \mathcal{H}$, we have $S' \subsetneq W_i$, then $T(S,W) = S$. However, if there exists $S' \in \mathcal{H}$ such that $S' \subseteq W_i$, then $T(S,W) = \emptyset$.

The construction method for a p-small cover relies on a randomized iterative process. Before elucidating the details of the process, we present the algorithm below. The overarching

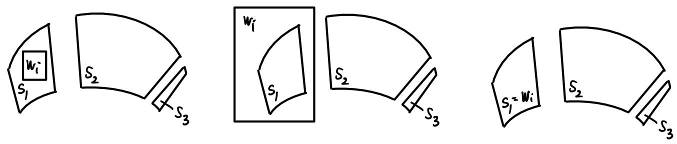

Figure 4.3: Edge Cases of Minimum $T(S, W_i)$ Fragment

strategy is to iteratively construct two sets of elements: $\bigcup_i \mathcal{U}_i$, $W = \bigcup_i W_i$. As we will see, depending on which stopping condition was enforced to interrupt the construction process, either $\bigcup_i \mathcal{U}_i$ is a p-small cover of $\langle \mathcal{H} \rangle$ or $W = \bigcup_i W_i$ possesses some edge of $langle \mathcal{H}\rangle$, which vindicates the overarching goal of showing that some uniformly random subset X_p^* of X admits a required property. This will be As Theorem 25 assumes that $\mathcal{H}$ is not p-small, the algorithm will output X_p^* w.h.p. .

The following definitions are needed: they will be recalled and justified later. Denote $l_i := 0.9^i l$, and let $\gamma := \lfloor \log_{0.9}(1/l) \rfloor + 1$ and $\zeta = \sqrt{\log_{0.9}(1/l)}$. We then define

$$L_i = \begin{cases} L, & \text{if } i < \gamma - \zeta \\ L\sqrt{\log l}, & \text{if } \gamma - \zeta \le i \le \gamma \end{cases}$$

Note the key assumption is that at the initialization of the process, the input hypergraph $\mathcal{H}_0$ is not p-small.

Algorithm 1: ConstructCover

Input : An l-bounded hypergraph $\mathcal{H}$ which is not p-small, a finite set X

Output: A family $\mathcal{U}$, a $m = (L_i p \log l)|X|$-sized random $W \in \binom{X}{m}$ subset of X.

$i := 0$, $X_0 := X$, $\mathcal{H}_0 := \mathcal{H}$, $\mathcal{U}_0 := \emptyset$, $\mathcal{G}_0 := \emptyset$;

while $\mathcal{H}_i \neq \{\emptyset\}$ *or* $\mathcal{H}_i \neq \emptyset$ **do**

  W_i = uniformly drawn from $\binom{X_i}{m}$;

  $X_{i+1} \longleftarrow X_i \setminus W_i$;

  $\mathcal{G}_i \longleftarrow \{S_{i-1} \in \mathcal{H}_{i-1} : t(S_{i-1}, W_i) \geq 0.9 l_i\}$;

  $\mathcal{U}_i \longleftarrow \{T(S_{i-1}, W_i) : S_{i-1} \in \mathcal{G}_i\}$;

  $\mathcal{H}_i \longleftarrow \{T(S_{i-1}, W_i) : S_{i-1} \in \mathcal{H}_{i-1} \setminus \mathcal{G}_i\}$;

end

return $\bigcup_i \mathcal{U}_i$, $W = \bigcup_i W_i$

Iteration Step: At each step i of the process, consider the host hypergraph $\mathcal{H}_{i-1}$ with hyperedges $\{S_i\}_{i=1}^{|\mathcal{H}_{i-1}|}$ obtained from the previous step, and the 'left over' set $X_i = X \setminus \cup_{j \leq i} W_j$. We sample an $w_i = L_i p|X|$-sized $W_i \in \binom{X_i}{w_i}$, and consider all induced minimum fragments $\{T(S_{i-1}, W_i) : S_{i-1} \in \mathcal{H}_{i-1}\}$. We generate the following sets:

$\boxed{1}$

$$\mathcal{G}_i = \{S_{i-1} \in \mathcal{H}_{i-1} : t(S_{i-1}, W_i) \geq 0.9l\},$$

a sub-hypergraph of $\mathcal{H}_{i-1}$ which contains elements S with 'large' minimum fragments with respect to W_i. In other words, the sets $S_{i-1} \in \mathcal{H}_{i-1}$ such that $|W_i \cap S_{i-1}| \leq 0.1 \cdot l$.

2
$$\mathcal{U}_i = \{T(S_{i-1}, W_i) : S \in \mathcal{G}_i\},$$
a cover for $\mathcal{G}_i$. Indeed, if the minimum fragment is large, we have $T(S_{i-1}, W_i) \subseteq S_{i-1} \in \mathcal{G}_i$.

3
$$\mathcal{H}_i = \{T(S_{i-1}, W_i) : S_{i-1} \in \mathcal{H}_{i-1} \setminus \mathcal{G}_i\},$$
the set of all minimum fragments with respect to W_i which are 'small'. This will be the host hypergraph in the next iteration step.

We note the following important characteristics of the sets generated at each step i. Firstly we have

$$\mathcal{H}_{i-1} \setminus \mathcal{G}_i \subseteq \langle \mathcal{H}_i \rangle, \tag{4.1}$$

and although $\mathcal{U}_i$ does not cover $\mathcal{H}_{i-1} \setminus \mathcal{G}_i$, it is the case that a cover of $\langle \mathcal{H}_i \rangle$ will cover $\mathcal{H}_{i-1} \setminus \mathcal{G}_i$. This reduces the task of finding a cover for this 'left over' set $\mathcal{H}_{i-1} \setminus \mathcal{G}_i$ to that of finding a cover for $\mathcal{H}_i$. We also note that

$$\text{for all } S_i \in \mathcal{H}_i, \text{ , we have } |S_i| \leq 0.9^i l, \tag{4.2}$$

as there exists $S_{i-1} \in \mathcal{H}_{i-1}$ such that $T(S_{i-1}, W_i) = S_i \in \mathcal{H}_i$, we have $|S_i| \leq 0.9^i l$, implying that $\mathcal{H}_i$ is $0.9^i l$-bounded. Finally,

$$\text{for all } S_i \in \mathcal{H}_i \text{ , we have } \left(\cup_{j \leq i} W_j \right) \cup S_i \in \langle \mathcal{H} \rangle, \tag{4.3}$$

a property which can be verified inductively. Indeed, Property 4.3 holds for an initial $W_0 = \emptyset$, and assuming it holds at step i, since each hyperedge $S_i \in H_i$ has the form $T(S_{i-1}, W_i) = S'_{i-1} \setminus W_i$ for some $S'_{i-1} \subseteq S_{i-1} \cup W_i$, we have

$$\left(\cup_{j \leq i} W_j \right) \cup S_i = \left(\cup_{j \leq i-1} W_j \right) \cup W_i \cup S'_{i-1} \setminus W_i \supseteq \left(\cup_{j \leq i-1} W_j \right) \cup W_i \cup S'_{i-1} \in \langle \mathcal{H} \rangle.$$

This property becomes important when considering the stopping conditions.

Stopping Condition: The process terminates when one of these two conditions holds.

1 Condition 1: $\mathcal{H}_i = \emptyset$
This will be the case if no minimum-fragment is small with respect to W_i. Proposition 38 shows that this entails $\mathcal{U} := \cup_{i \leq \gamma} \mathcal{U}_i$ covers $\mathcal{H}$.

2 Condition 2: $\mathcal{H}_i = \{\emptyset\}$
This will be the case if all minimum fragments which are small with respect to W_i are sets of a single element. In this case, Proposition 38 shows that this entails $W = \cup_i W_i \in \langle \mathcal{H} \rangle$. In particular, $\mathcal{U} := \cup_{i \leq \gamma} \mathcal{U}_i$ does not necessarily cover $\mathcal{H}$.

Condition 2 is equivalent to $\forall S_i \in \mathcal{H}_i$, $|S_i| < 1$. This implies that the step γ in the process such that $\forall S_i \in \mathcal{H}_i$, $|S_i| < 1$. Taking a fragment of the largest set $S_{i-1} \in \mathcal{H}_{i-1}$ would result in a set of size less than one. Formally, γ is the largest i such that $\forall S_i \in \mathcal{H}_i$, $\lfloor 0.9 |S_i| \rfloor \geq 1$, or $\gamma = \lfloor \log_{0.9}(1/l) \rfloor + 1$. Condition 2 says that the process terminates in at most γ steps.

Proposition 38. For $\gamma > \log_{0.9}(1/l)$, either $W \in \langle \mathcal{H} \rangle$ or $\mathcal{U} := \cup_{i \leq \gamma} \mathcal{U}_i$ covers $\mathcal{H}$.

Proof. The stopping conditions tell us that for $\gamma > \log_{0.9}(1/l)$, either $\mathcal{H}_i = \{\emptyset\}$ or $\mathcal{H}_i = \emptyset$.
If the process terminates due to Condition 2, then we have $\mathcal{H}_\gamma = \{\emptyset\}$. Property 4.3 says that $W = (\cup_{i \leq \gamma}) W_i \cup \emptyset \in \langle \mathcal{H} \rangle$. That is, the $Lp \log l$-uniformly random subset W_i of X belongs to $\langle \mathcal{H} \rangle$.
On the other hand if the process terminates due to Condition 1, $\mathcal{H}_\gamma = \emptyset$. At each step i of the process, for any $S_i = T(S_{i-1}, W_i)$ we have either $S_i \in \mathcal{G}_i$ and hence $\mathcal{U}_i$, or we define $S_{i+1} = S_i \in \mathcal{H}_{i+1}$. Since $\mathcal{H}_\gamma = \emptyset$, there is $j < \gamma$ for which $S_j \in \mathcal{G}_j$. So $S = S_1 \cup S_2 \dots S_j$ is covered by $\cup_{i \leq j} \mathcal{U}_i$, and hence S is covered by $\mathcal{U}$. ◻

Denote by Φ and Ψ respectively the events "$\mathcal{U}$ covers $\mathcal{H}$" and "$W \in \langle \mathcal{H} \rangle$". By Proposition 38, we have $\mathbb{P}(\Phi) + \mathbb{P}(\Psi) \geq 1$, and rearranging, we obtain

$$\mathbb{P}(\Psi) \geq 1 - \mathbb{P}(\Phi).$$

Recall that proving the reformulated conjecture Theorem 32 amounts to showing that $\mathbb{P}(\Psi) = 1 - o_{l \to \infty}(1)$. In the next section we present the key argument in [park,pham], namely that $\mathbb{P}(\Phi) = o_{l \to \infty}(1)$.

4.3 Proof of the Reformulated Conjecture

In this section we cover the second cmponent of the proof of Theorem 32: we use $\mathcal{U} = \cup_i \mathcal{U}_i$ and $W = \cup_i W_i$ constructed in the previous section to show that W belongs to $\langle \mathcal{H} \rangle$ with probability $1 - o(1)$.

The assumption that $\mathcal{H}$ is not p-small implies

$$\text{if } \Phi \text{ occurs, then } \sum_{U \in \mathcal{U}(W)} p^{|U|} > 1/2,$$

and applying Markov's inequality provides the following initial bound, we obtain

$$\mathbb{P}(\Phi) \leq \mathbb{P}\left(\sum_{U \in \mathcal{U}(W)} p^{|U|} > 1/2 \right) \overset{\text{Lemma 10}}{\leq} 2\mathbb{E}\left[\sum_{U \in \mathcal{U}(W)} p^{|U|} \right]. \tag{4.4}$$

We now introduce the central lemma to the argument: bounding the expected cost of $\mathcal{U}$.

Lemma 39. *For $W_i \in \binom{X}{w_i}$ a uniformly chosen w-subset of X, and $l_i := 0.9^i l$ we have*

$$\mathbb{E}\left[\sum_{U \in \mathcal{U}_i} p^{|U|} \right] < L^{-0.8 l_i}.$$

As the expectation is over the randomness of W_i, we expand the expectation term and obtain

$$\mathbb{E}\left[\sum_{U \in \mathcal{U}_i} p^{|U|} \right] = \binom{n}{w_i}^{-1} \sum_{W_i \in \binom{X}{w_i}} \sum_{U \in \mathcal{U}_i} p^{|U|} < (\binom{n}{w_i})^{-1} \binom{n}{w_i} L^{-0.8 l_i}.$$

Hence Lemma 39 is equivalent to the following proposition.

Proposition 40. For $W_i \in \binom{X}{w_i}$ a uniformly chosen w_i-subset of X, with $|X| = n$, we have

$$\sum_{W_i \in \binom{X}{w_i}} \sum_{U \in \mathcal{U}_i} p^{|U|} < \binom{n}{w} L^{-0.8 l_i}.$$

Proof. Denote at each step i,

$$\mathcal{G}_{m_i} := \{S_{i-1} \in \mathcal{H}_{i-1} : t(S_{i-1}, W_i) = m_i\}, \text{ and note that } \mathcal{G}_i := \bigcup_{m_i = \lfloor l_i \rfloor + 1}^{\lfloor l_{i-1}} \mathcal{G}_{m_i}.$$

Similarly, let

$$\mathcal{U}_{m_i} := \{T(S_{i-1}, W_i) : S_{i-1} \in \mathcal{G}_{m_i}\}, \text{ and note that } \mathcal{U}_i := \bigcup_{m_i = \lfloor l_i \rfloor + 1}^{\lfloor l_{i-1}} \mathcal{U}_{m_i}.$$

$$\mathcal{U}(W) = \bigcup_{i \le \gamma} \mathcal{U}_i = \bigcup_{i \le \gamma} \left(\bigcup_{m_i = \lfloor l_i \rfloor + 1}^{\lfloor l_{i-1} \rfloor - 1} \mathcal{U}_{m_i} \right), \tag{4.5}$$

where the unions on the right hand side are disjoint. In addition, for any $U \in \mathcal{U}_{m_i}$, there is $S_{i-1} \in \mathcal{H}_{i-1}$ such that $|U| = t(S_{i-1}, W_i) = m_i$. Hence for a given m_i, the right hand side of 40 is a sum over all m_i-sized minimal fragments with respect to W_i, and over all possible W_i

$$\sum_{W_i \in \binom{W_i}{w_i}} \sum_{U \in \mathcal{U}_{m_i}} p^{|U|} = p^{m_i} \left| \left\{ (W_i, T(S_{i-1}, W_i)) : W_i \in \binom{W_i}{w_i}, S \in \mathcal{H} \text{ , and } t(S_{i-1}, W_i) = m_i \right\} \right|.$$

Here the notion of a minimum fragment becomes key to bound the number of choices of $(W_i, T_i := T(S_{i-1}, W_i))$. The following steps lead up to this bound.

Step 1 Letting $Z_i = W_i \cup T_i$, and noting that by definition $W \cap T = \emptyset$. Hence $|Z| = w_i + m_i$. Hence the number of choices for (W_i, T), recalling that $w_i = Lpn$ is

$$\binom{n}{w_i + m_i} = \binom{n}{m_i} \prod_{j=0}^{m_i - 1} \frac{n - m_i - j}{m_i + j + 1} \le \binom{n}{m_i} \prod_{j=0}^{m_i - 1} \frac{n}{m_i} \le \binom{n}{m_i} \prod_{j=0}^{m_i - 1} (Lp)^{-m_i}$$

Step 2 An obvious upper bound on the number of choices of $T \subset Z$ is $2^{w_i + m_i}$. The choice of (W_i, T_i) is then determined upon fixing a choice of Z_i and T_i. But we can do better. Z_i must contain an edge of $\mathcal{H}$ by definition of minimum fragment, so there must exist some $\tilde{S}_{i-1} \in \mathcal{H}_{i-1}$ such that $\tilde{S}_{i-1} \subseteq Z$. Pick one such $\tilde{S}_{i-1}$ arbitrarily. From Proposition 37, we note that $T \subseteq \tilde{S}_{i-1}$. This allows the specification of T as a subset of $\tilde{S}_{i-1}$, the number of possibilities of which is at most 2^{l_i}. This yields

$$\sum_{W_i \in \binom{W_i}{w_i}} \sum_{U \in \mathcal{U}_{m_i}} p^{|U|} \le p^{m_i} \binom{n}{w_i} (Lp)^{-m_i} 2^{l_i} = \binom{n}{w_i} L^{-m_i} 2^{l_i},$$

and simplifying by the number of choices of W_i we obtain

$$\sum_{U\in\mathcal{U}_{m_i}} p^{|U|} \leq p^{m_i}(Lp)^{-m_i}2^{l_i} = L^{-m_i}2^{l_i}. \tag{4.6}$$

From the Equation 4.5, bounding the left hand side of Proposition 40 at each step $i \geq 1$, and for

$$L \geq \max\{2, (0.1l \cdot 2^l)^{0.8l+\lfloor l\rfloor-1}\}, \tag{4.7}$$

we obtain

$$\sum_{U\in\mathcal{U}_i} p^{|U|} \overset{(4.5)}{=} \sum_{m_i=\lfloor l_i\rfloor+1}^{\lfloor l_{i-1}\rfloor-1} \sum_{U\in\mathcal{U}_{m_i}} p^{|U|} \overset{(4.4)}{\leq} \sum_{m_i=\lfloor l_i\rfloor+1}^{\lfloor l_{i-1}\rfloor-1} L^{-m_i}2^{l_i} \leq L^{-0.8l_i}. \tag{4.8}$$

as for each step i, one has $l_i < l$ and hence $L \geq (0.1l_i \cdot 2^{l_i})^{0.8l_i+\lfloor l_i\rfloor-1}$. With $\lfloor l_{i-1}\rfloor - 1 - (\lfloor l_i\rfloor + 1) \leq l_{i-1} - l_i = 0.9^{i-1}(1-0.9)l \leq 0.1l$ we have

$$\sum_{m_i=\lfloor l_i\rfloor+1}^{\lfloor l_{i-1}\rfloor-1} L^{-m_i}2^{l_i} \leq 0.1l_i \cdot L^{-(\lfloor l_i\rfloor+1)}2^{l_i} \leq L^{0.8l_i}.$$

□

This verifies Proposition 40. We can now use Lemma 39 to continue to bound $\mathbb{P}(\Phi)$. To do so, we wish to specify the upper bound on the expected cost of the cover $\mathcal{U}(W)$ in Property 4.4. Recalling from Equation 4.5 that the covers $\mathcal{U}_i$ created at each step i of the process are disjoint, we have

$$\mathbb{E}\left[\sum_{U\in\mathcal{U}(W)} p^{|U|}\right] \overset{4.5}{=} \sum_{i\leq\gamma}\mathbb{E}\left[\sum_{U\in\mathcal{U}_i} p^{|U|}\right] \overset{4.8}{\leq} \sum_{i\leq\gamma} L^{-0.8l_i}.$$

We now show how we can obtain a bound of order $o_{l\to\infty}(1)$ on the expected cost of the cover $U(W) = \cup_{i\leq\gamma}\mathcal{U}_i$. In particular, for the purpose of faster convergence, we will choose slightly larger random subsets towards the last iterations of the cover creation process. Recall that $W = \cup_{i\leq\gamma}W_i$ resulting from Algorithm 1 is a uniformly random subset of X of size $O(Lp\log l|X|)$. Recall $\zeta := \sqrt{\log_{0.9}(1/l)}$, and

$$L_i = \begin{cases} L, & \text{if } i < \gamma-\zeta \\ L\sqrt{\log l}, & \text{if } \gamma-\zeta \leq i \leq \gamma \end{cases}$$

Our goal is to estimate $\mathbb{E}\left[\sum_{U\in\mathcal{U}(W)} p^{|U|}\right]$. Note that

$$\mathbb{E}\left[\sum_{U\in\mathcal{U}(W)} p^{|U|}\right] \leq \sum_{1\leq i\leq\gamma} L_i^{-0.8l_i} = \sum_{1\leq i<\gamma-\zeta} L^{-0.8l_i} + \sum_{\gamma-\zeta\leq i\leq\gamma} L_i^{-0.8l_i}. \tag{4.9}$$

Proposition 41. With $W = \cup_{i\leq\gamma}W_i$ and $U(W) = \cup_{i\leq\gamma}\mathcal{U}_i$ resulting from Algorithm 1, we have

$$\sum_{1\leq i<\gamma-\zeta} L_i^{-0.8l_i} \leq 2L^{-0.8\exp(c_1\sqrt{\log l})} \text{ and } \sum_{\gamma-\zeta\leq i<\gamma} L_i^{-0.8l_i} = O\left((L\sqrt{\log l})^{c_2}\right)$$

for some $c_1, c_2 > 0$.

Proof. **Iterations up to step $\gamma - \zeta$**
In these iterations, we sample uniformly random subsets of X of size Lpn. We first note that for $i < \gamma - \zeta$, there exists an absolute constant $c_1 > 0$ such that $l_i \geq \exp(c_1\sqrt{\log l})$. In particular, if for each $q \in \mathbb{N}$ such that $q < \gamma - \zeta$, we have $\sum_{i \leq q} L^{-0.8l_i} \leq 2L^{-0.8l_q}$, then we will obtain $\sum_{i<q} L^{-0.8l_i} < 2L^{\exp(c_1\sqrt{\log l})}$. Let us show the former by induction. Assuming without loss of generality that $\zeta > 1$, we confirm that for $q = 1$ we indeed have $L^{-0.8 \cdot 0.9 l_1} \leq 2L^{-0.8 \cdot 0.9 l_1}$. Suppose that $\sum_{i \leq q} L^{-0.8l_i} \leq 2L^{-0.8l_q}$ holds for some $q < \zeta$. We have, for $q' = q + 1$,

$$\sum_{i \leq q'} L^{-0.8l_i} \leq \sum_{i \leq q} L^{-0.8l_i} + L^{-0.8l_{q'}} \leq 2L^{-0.8l_q} + \left(L^{-0.8l_q}\right)^{0.9} \overset{(*)}{\leq} 2\left(L^{-0.8l_q}\right)^{0.9}$$

where $2L^{-0.8l_q} < \left(L^{-0.8l_q}\right)^{0.9}$ for some L large enough. Indeed, let L be such that Property 4.7 is satisfied, and in addition $L > 2^{-0.08l_q}$. Indeed, after some rearranging, this is equivalent to $2L^{-0.8l_q} < \left(L^{-0.8l_q}\right)^{0.9}$. We can now conclude

$$\sum_{1 \leq i < \gamma - \zeta} L^{-0.8l_i} \leq 2L^{-0.8l_{\lfloor \gamma - \zeta \rfloor}} \leq 2L^{-0.8\exp(c_1\sqrt{\log l})}.$$

Last iterations
These iterations of the algorithm are characterized by larger random samples $W_i \in \binom{X_i}{w_i}$ because $w_i = L_i pn = (L \log l)pn$. The terms of the following finite sum are decreasing in i and in L_i. In particular, larger samples allow faster convergence of the bounds in the last iterations.

$$\begin{aligned}\sum_{\gamma - \zeta \leq i < \gamma} L_i^{-0.8l_i} &= \sum_{\gamma - \zeta \leq i < \gamma} (L\sqrt{\log l})^{-0.8l_i} \leq \lfloor \zeta \rfloor (L\sqrt{\log l})^{\lfloor \gamma - \zeta \rfloor + 1} \\ &= O\big((L\sqrt{\log l})^{c_2^*}\big) = O\big((L \log l)^{c_2}\big)\end{aligned}$$

for some $c_2 > 2c_2^* > 0$. We are now ready to prove the bound on the cost of the cover $\mathcal{U}(W)$. □

Lemma 42. *With $W = \cup_{i \leq \gamma} W_i$ and $\mathcal{U}(W) = \cup_{i \leq \gamma} \mathcal{U}_i$ resulting from Algorithm 1, the cost of the cover $\mathcal{U}(W)$ is*

$$\mathbb{E}\left[\sum_{U \in \mathcal{U}(W)} p^{|U|}\right] \leq (\log l)^{-c},$$

for some $c > 0$.

Proof. Combining the results obtained in Property 4.9 and Lemma 41, we have

$$\mathbb{E}\left[\sum_{U \in \mathcal{U}(W)} p^{|U|}\right] \leq \sum_{1 \leq i < \gamma - \zeta} L^{-0.8l_i} + \sum_{\gamma - \zeta \leq i \leq \gamma} L_i^{-0.8l_i} \leq 2L^{-0.8\exp(c_1\sqrt{\log l})} + O\big((L \log l)^{c_2}\big)$$

Noting that one can pick $c_3 > 0$ such that $2L^{-0.8\exp(c_1\sqrt{\log l})} + O\big((L \log l)^{c_2}\big) = (\log l)^{-c_3}$, we obtain

$$\mathbb{E}\left[\sum_{U \in \mathcal{U}(W)} p^{|U|}\right] \leq (\log l)^{-c_3} \tag{4.10}$$

□

Finalizing our upper bound on the expected cost of the cover $\mathcal{U}(W)$ allows us to bound the probability of event that $\mathcal{U}(W)$ covers $\mathcal{H}$ as discussed.

$$\mathbb{P}(\Phi) \leq \mathbb{P}\left(\sum_{U \in \mathcal{U}(W)} p^{|U|} > 1/2\right) \overset{\text{Lemma 10}}{\leq} 2\mathbb{E}\left[\sum_{U \in \mathcal{U}(W)} p^{|U|}\right] \overset{(4.10)}{\leq} (\log l)^{-c_3} = o_{l \to \infty}(1). \tag{4.11}$$

From the exhaustivity of the two possibilities from the output of Algorithm 1 outlined in Proposition 38, we obtain $\mathbb{P}(\Psi) \geq 1 - \mathbb{P}(\Phi) = 1 - o_{l \to \infty}(1)$. In other words, $W \in \mathcal{H}$ w.h.p.. This verifies Theorem 32, and by Proposition 35 also verifies Theorem 25, the original formulation of the Kahn-Kalai Conjecture.

Bibliography

Alweiss, R., S. Lovett, K. Wu, and J. Zhang (2021). Improved bounds for the sunflower lemma. *Ann. of Math. (2)* 194(3), 795–815. URL https://doi.org/10.4007/annals.2021.194.3.5.

Bell, T. and A. Frieze (2024). Rainbow powers of a Hamilton cycle in $G_{n,p}$. *J. Graph Theory* 105(4), 491–500.

Bollobás, B., S. Janson, and O. Riordan (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* 31(1), 3–122. URL https://doi.org/10.1002/rsa.20168.

Bollobás, B. and A. Thomason (1987). Threshold functions. *Combinatorica* 7(1), 35–38. URL https://doi.org/10.1007/BF02579198.

Diestel, R. (2017). *Graph Theory*. Springer Publishing Company, Incorporated, 5th edition.

Erdős, P. and A. Rényi (1960). On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 5, 17–61.

Fischer, M., N. Škorić, A. Steger, and M. Trujić (2022). Triangle resilience of the square of a Hamilton cycle in random graphs. *J. Combin. Theory Ser. B* 152, 171–220. URL https://doi.org/10.1016/j.jctb.2021.09.005.

Frankston, K., J. Kahn, B. Narayanan, and J. Park (2021). Thresholds versus fractional expectation-thresholds. *Ann. of Math. (2)* 194(2), 475–495. URL https://doi.org/10.4007/annals.2021.194.2.2.

Friedgut, E. (2005). Hunting for sharp thresholds. *Random Structures Algorithms* 26(1-2), 37–51. URL https://doi.org/10.1002/rsa.20042.

Frieze, A. and M. Karoński (2015). *Introduction to Random Graphs*. Cambridge University Press.

Johansson, A., J. Kahn, and V. Vu (2008). Factors in random graphs. *Random Structures Algorithms* 33(1), 1–28. URL https://doi.org/10.1002/rsa.20224.

Kahn, J. and G. Kalai (2007). Thresholds and expectation thresholds. *Combin. Probab. Comput.* 16(3), 495–502. URL https://doi.org/10.1017/S0963548307008474.

Kahn, J., B. Narayanan, and J. Park (2021). The threshold for the square of a Hamilton cycle. *Proc. Amer. Math. Soc.* 149(8), 3201–3208. URL https://doi.org/10.1090/proc/15419.

Klenke, A. (2020). *Probability theory—a comprehensive course*. Universitext. Springer, Cham, third edition. URL https://doi.org/10.1007/978-3-030-56402-5.

Kühn, D. and D. Osthus (2012). On Pósa's conjecture for random graphs. *SIAM J. Discrete Math.* 26(3), 1440–1457. URL https://doi.org/10.1137/120871729.

Mitzenmacher, M. and E. Upfal (2005). *Probability and computing*. Cambridge University Press, Cambridge. URL https://doi.org/10.1017/CBO9780511813603, randomized algorithms and probabilistic analysis.

Montgomery, R. (2018). Topics in Random Graphs. https://www.math.kit.edu/iag6/lehre/semprobmeth2020s/media/topics-in-random-graphs.pdf.

Park, J. and H. T. Pham (2022). A proof of the Kahn-Kalai conjecture. In: *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science—FOCS 2022*, 636–639. IEEE Computer Soc., Los Alamitos, CA.

Pósa, L. (1976). Hamiltonian circuits in random graphs. *Discrete Math.* 14(4), 359–364. URL https://doi.org/10.1016/0012-365X(76)90068-6.

Rao, A. (2023). The sunflower lemma and monotone thresholds. URL https://www.youtube.com/watch?v=Rw6FEZnuE5Y.

Riordan, O. (2000). Spanning subgraphs of random graphs. *Combin. Probab. Comput.* 9(2), 125–148. URL https://doi.org/10.1017/S0963548399004150.

Skerman, F. (2008). Lecture 2: Threshold for a random graph containing a cycle. https://fskerman.github.io/Lect2.pdf.

van der Hofstad, R. (2017). *Random graphs and complex networks. Vol. 1*, volume [43] of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge. URL https://doi.org/10.1017/9781316779422.